

INTRODUCTORY VERSION

Database Systems

An Application-Oriented Approach

SECOND EDITION



Michael KIFER Arthur BERNSTEIN Philip M. LEWIS

TP311.13
K47
E-2

INTRODUCTORY VERSION

Database Systems

An Application-Oriented Approach

SECOND EDITION

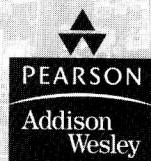


Michael Kifer Arthur Bernstein Philip M. Lewis

STATE UNIVERSITY OF NEW YORK, STONY BROOK



E200500271



Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

<i>Executive Editor</i>	Susan Hartman Sullivan
<i>Acquisitions Editor</i>	Matt Goldstein
<i>Project Editor</i>	Maite Suarez-Rivas
<i>Marketing Manager</i>	Nathan Schultz
<i>Senior Marketing Coordinator</i>	Lesly Hershman
<i>Senior Production Supervisor</i>	Jeffrey Holcomb
<i>Project Management</i>	Windfall Software
<i>Copyeditor</i>	Elisabeth Beller
<i>Proofreader</i>	Jennifer McClain
<i>Composition</i>	Windfall Software, using ZzT _E X
<i>Text Designer</i>	Sandra Rigney
<i>Cover Designer</i>	Joyce Cosentino Wells
<i>Cover Photo</i>	© 2004 Digital Vision
<i>Prepress and Manufacturing</i>	Caroline Fell

Access the latest information about Addison-Wesley titles from our World Wide Web site:
<http://www.aw-bc.com/computing>

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial caps or all caps.

The programs and applications presented in this book have been included for their instructional value. They have been tested with care, but are not guaranteed for any particular purpose. The publisher does not offer any warranties or representations, nor does it accept any liabilities with respect to the programs or applications.

Library of Congress Cataloging-in-Publication Data

Kifer, M. (Michael), 1954–

Database systems : an application-oriented approach / Michael Kifer, Arthur Bernstein,
Philip M. Lewis.—Introductory version, 2nd ed.

p. cm.

Lewis's name appears first on the earlier edition.

Rev. ed. of: Databases and transaction processing / Philip M. Lewis. 2002.

ISBN 0-321-22838-3 (hardcover)

1. Database management. 2. Transaction systems (Computer systems) I. Bernstein, Arthur J.
II. Lewis, Philip M., 1931– Databases and transaction processing. III. Title.

QA76.9.D3K4965 2004

005.74—dc22

2003068925

Copyright © 2005 by Pearson Education, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ISBN 0-321-22838-3

1 2 3 4 5 6 7 8 9 10—CRW—08 07 06 05

INTRODUCTORY VERSION

Database Systems

An Application-Oriented Approach

SECOND EDITION

*In memory of my late parents, Luba and Isaac;
and to my wife, Lora, and my children. M.K.*

To my wife, Edie, my children, and my grandchildren. A.J.B.

To my wife, Rhoda, my children, and my grandchildren. P.M.L.

Preface

We are publishing the second edition of our textbook in two versions:

- This version, which consists of introductory material, is appropriate for a first undergraduate or graduate course in databases.
- The second version, which is the complete book, is appropriate for three courses:
 - An introductory undergraduate or graduate course in databases
 - An undergraduate or graduate course in transaction processing for students who have had an introductory course in databases
 - An advanced undergraduate or a first graduate course in databases for students who have had an introductory course in databases

One of our goals was to reduce the size and make this introductory version more affordable to students. Another was to capitalize on our experience in using the first edition of the book to make an even better introductory text.

The chapters in this book are not just a subset of those in the complete book. We believe that instructors of an introductory database course should have the option of enriching an introductory course by including material on object databases and XML—topics that are covered in great detail in several chapters in the complete book. Therefore we have added to the introductory book two new chapters, Chapter 16, Introduction to Object Databases, and Chapter 17, Introduction to XML and Web Data, which contain an appropriately chosen subset of the material in the full version of this book.

To keep the book up-to-date with the rapidly changing technology, we have added a substantial amount of material on UML to a number of chapters and have included a new chapter on Database Tuning, Chapter 12, in both the introductory and complete books.

As with the first edition, our focus is on how to build applications using databases rather than on how to build the database management system itself. We believe that many more students will be implementing applications than will be building DBMSs. Thus, we include substantial material describing the languages and APIs used by transactions to access a database, such as embedded SQL, ODBC, and JDBC.

Although we cover many practical aspects of database and transaction processing applications, we are primarily concerned with the concepts that underlie these topics rather than with the details of particular commercial systems or applications.

Thus we concentrate on the concepts behind the relational and object data models. These concepts will remain the foundation of database processing long after SQL is obsolete.

To enhance students' understanding of the technical material, we have included a case study of a transaction processing application, the Student Registration System, which is carried through the book. While a student registration system can hardly be considered glamorous, it has the unique advantage that all students have interacted with such a system as users. More importantly, it turns out to be a surprisingly rich application, so we can use it to illustrate many of the issues in database design, query processing, and transaction processing.

A unique aspect of the book is a presentation of the software engineering concepts required to implement transaction processing applications, using the Student Registration System as an example. Since the implementations of many information systems fail because of poor project management and inadequate software engineering, we feel that these topics should be an important part of the student's education. Our treatment of software engineering issues is brief, since many students will take a separate course in this subject. However, we believe that they will be better able to understand and apply that material when they see it presented in the context of an information system implementation. Since the courses that use this text at Stony Brook are not software engineering courses, we do not cover this material in class. Instead, we ask the students to read it and require that they use good software engineering practice in their class projects. We do cover in class those aspects of the Student Registration System that illustrate important issues in databases and transaction processing.

Changes in the Second Edition

The technology underlying database and transaction processing systems is changing so rapidly that we have made a large number of changes and additions to the material of the first edition. One rapidly advancing technology is the Unified Modeling Language, UML. We added substantial amount of material on UML in Chapter 4 on database design, in addition to the material on E-R diagrams that was already there. We also added UML to the material on software engineering in Chapters 2, 14, and 15.

A new chapter on Database Tuning, Chapter 12, was added because so much effort in the real world is spent increasing the throughput of database and transaction processing applications.

In addition, material has been added and updated in almost all the chapters. Significant examples of this are the coverage of SQL/XML and RAID technology.

One important area that is *not* included in this volume is Web Services. Since this is a rapidly developing and interesting application-oriented subject we have significantly revised the complete version of this text to include material on this topic. In addition to strengthening the book on the subject of XML Technology by updating the chapter on XML and Web Data and adding a section on SQL/XML, we have added a new chapter on Web Services that contains material on SOAP,

WSDL, BPML, UDDI, and XML-based transaction processing using WS-Coordination and WS-Transaction. In the chapter on Security and Internet Commerce, we added a section on XML-based encryption, using XML-Encryption, XML-Signature, WS-Security, and SAML. And in the chapter on Architecture of Transaction Processing Systems, we added material on Web Application Servers and J2EE, which are used to implement the back-end of many Web services.

Organization of the Book

Chapters 1 through 7 should be taught in the order in which they appear in the book. Chapter 8 contains much of the information that students need in order to put the knowledge they acquired in the preceding chapters into practice. However, subsequent chapters do not significantly depend on Chapter 8. Chapters 9 through 12 in Part 3 should be taught sequentially. Chapter 13 in the same part is largely independent. The software engineering chapters in Part 4 utilize the material of the chapters in Parts 2 and 3, but the software engineering chapters can be read in parallel with the database material. Chapters 16 and 17 in the advanced part of the book depend on the first seven chapters in Part 2.

Finally we note that the sections in this book that are marked with an asterisk (*) are optional and can be omitted, if the instructor prefers to do so. Sections marked with the © icon in the table of contents deal with the case study. Also, exercises that are marked with an asterisk are slightly harder than the rest, and exercises that are marked with two asterisks are even harder.

Supplements

In addition to the text, the following supplementary materials are available to assist instructors:

- Online PowerPoint presentations for all chapters
- Online PowerPoint slides of all figures
- An online solution manual containing solutions for the exercises
- Additional references, notes, errata, homeworks, and exams.

For more information on obtaining these supplements, please visit this book's Companion Website at www.aw-bc.com/kifer. The solutions manual and PowerPoint presentations are available only to instructors through your Addison-Wesley sales representative. To contact your representative, please visit www.aw.com.

Acknowledgments

We would like to thank the reviewers, whose comments and suggestions significantly improved the second edition of the book:

Tran Cao Son, New Mexico State University
Frantisek Franek, McMaster University

Junping Sun, Nova Southeastern University
Philip Cannata, Sun Microsystems
Dehu Qi, Lamar University
Nematollah Shiri, Concordia University
Jian Pei, State University of New York at Buffalo
Jack Wileden, University of Massachusetts Amherst
Sibel Adali, Rensselaer Polytechnic Institute
Roger King, University of Colorado at Boulder
Markus Schneider, University of Florida
Yaron Y. Goland, BEA
Dennis Shasha, New York University
Christelle Scharff, Pace University
Zhiwei Wang, Graduate Programs in Software Engineering, IT, and IS University
of St. Thomas.

We would also like to thank the reviewers of the first edition of the book:

Suad Alagic, Wichita University
Catriel Beeri, The Hebrew University
Rick Cattell, Sun Microsystems
Jan Chomicki, SUNY Buffalo
Henry A. Etlinger, Rochester Institute of Technology
Leonidas Fegaras, University of Texas at Arlington
Alan Fekete, University of Sydney
Johannes Gehrke, Cornell University
Hershel Gottesman, Consultant
Jiawei Han, Simon Fraser University
Peter Honeyman, University of Michigan
Vijay Kumar, University of Missouri–Kansas City
Jonathan Lazar, Towson University
Dennis McLeod, University of Southern California
Rokia Missaoui, University of Quebec in Montreal
Clifford Neuman, University of Southern California
Fabian Pascal, Consultant
Sudha Ram, University of Arizona
Krithi Ramamritham, University of Massachusetts–Amherst, and IIT Bombay
Andreas Reuter, International University in Germany, Bruchsal
Arijit Sengupta, Georgia State University

Munindar P. Singh, North Carolina State University

Greg Speegle, Baylor University

Junping Sun, Nova Southeastern University

Joe Trubisz, Consultant

Vassilis J. Tsotras, University of California, Riverside

Emilia E. Villarreal, California Polytechnic State University

We would also like to thank the following people who were kind enough to provide us with additional information and answers to our questions: Don Chamberlin, Daniela Florescu, Jim Gray, Pankaj Gupta, Rob Kelly, and C. Mohan.

Two people taught out of beta versions of the book and made useful comments and suggestions: David S. Warren and Radu Grosu. Joe Trubicz served not only as a reviewer when the manuscript was complete, but provided critical comments on early versions of many of the chapters.

A number of students were very helpful in reading and checking the correctness of various parts of the book: Ziyang Duan, Shiyong Lu, Swapnil Patil, Guizhen Yang, and Yan Zhang.

Many thanks to the staff of the Computer Science Department at Stony Brook, and in particular Kathy Germana, who helped make things happen at work.

We would particularly like to thank Matt Goldstein and Maite Suarez-Rivas, our editors at Addison-Wesley, who played an important role in shaping the contents and approach of the book in its early stages and throughout the time we were writing it. We would also like to thank the various staff members of Addison-Wesley and Windfall Software, who did an excellent job of editing and producing the book: Jeffrey Holcomb, Paul Anagnostopoulos, Elisabeth Beller, Jennifer McClain, and Joe Snowden.

Last, but not least, we would like to thank our wives, Lora, Edie, and Rhoda, who provided much needed support and encouragement while we were writing the book.

Contents

Preface	xvii
PART ONE Introduction	1
1 Overview of Databases and Transactions	3
1.1 What Are Databases and Transactions?	3
1.2 Features of Modern Database and Transaction Processing Systems	6
1.3 Major Players in the Implementation and Support of Database and Transaction Processing Systems	7
1.4 Decision Support Systems—OLAP and OLTP	9
2 The Big Picture	13
⑈ 2.1 Case Study: A Student Registration System	13
2.2 Introduction to Relational Databases	14
2.3 What Makes a Program a Transaction—The ACID Properties	20
Bibliographic Notes	25
Exercises	25
PART TWO Database Management	29
3 The Relational Data Model	31
3.1 What Is a Data Model?	31
3.2 The Relational Model	35
3.2.1 Basic Concepts	35
3.2.2 Integrity Constraints	38
3.3 SQL—Data Definition Sublanguage	46
3.3.1 Specifying the Relation Type	46
3.3.2 The System Catalog	46
3.3.3 Key Constraints	47
3.3.4 Dealing with Missing Information	48

3.3.5 Semantic Constraints	49
3.3.6 User-Defined Domains	53
3.3.7 Foreign-Key Constraints	53
3.3.8 Reactive Constraints	56
3.3.9 Database Views	59
3.3.10 Modifying Existing Definitions	60
3.3.11 SQL-Schemas	62
3.3.12 Access Control	63
Bibliographic Notes	65
Exercises	66

4 Conceptual Modeling of Databases with Entity-Relationship Diagrams and the Unified Modeling Language **69**

4.1 Conceptual Modeling with the E-R Approach	70
4.2 Entities and Entity Types	70
4.3 Relationships and Relationship Types	73
4.4 Advanced Features in Conceptual Data Modeling	78
4.4.1 Entity Type Hierarchies	78
4.4.2 Participation Constraints	81
4.4.3 The Part-of Relationship	83
4.5 From E-R Diagrams to Relational Database Schemas	86
4.5.1 Representation of Entities	86
4.5.2 Representation of Relationships	88
4.5.3 Representing IsA Hierarchies in the Relational Model	90
4.5.4 Representation of Participation Constraints	92
4.5.5 Representation of the Part-of Relationship	94
4.6 UML: A New Kid on the Block*	95
4.6.1 Representing Entities in UML	96
4.6.2 Representing Relationships in UML	97
4.6.3 Advanced Modeling Concepts in UML	101
4.6.4 Translation to SQL	105
4.7 A Brokerage Firm Example	106
4.7.1 An Entity-Relationship Design	106
4.7.2 A UML Design*	110
④ 4.8 Case Study: A Database Design for the Student Registration System	111
4.8.1 The Database Part of the Requirements Document	112
4.8.2 The Database Design	113
4.9 Limitations of Data Modeling Methodologies	119
Bibliographic Notes	123
Exercises	123

5	Relational Algebra and SQL	127
5.1	Relational Algebra: Under the Hood of SQL	128
5.1.1	Basic Operators	128
5.1.2	Derived Operators	137
5.2	The Query Sublanguage of SQL	147
5.2.1	Simple SQL Queries	148
5.2.2	Set Operations	154
5.2.3	Nested Queries	157
5.2.4	Quantified Predicates	163
5.2.5	Aggregation over Data	164
5.2.6	Join Expressions in the FROM Clause	170
5.2.7	A Simple Query Evaluation Algorithm	171
5.2.8	More on Views in SQL	174
5.2.9	Materialized Views	177
5.2.10	The Null Value Quandary	181
5.3	Modifying Relation Instances in SQL	182
5.3.1	Inserting Data	182
5.3.2	Deleting Data	184
5.3.3	Updating Existing Data	185
5.3.4	Updates on Views	185
	Bibliographic Notes	187
	Exercises	188
6	Database Design with the Relational Normalization Theory	193
6.1	The Problem of Redundancy	193
6.2	Decompositions	195
6.3	Functional Dependencies	198
6.4	Properties of Functional Dependencies	200
6.5	Normal Forms	207
6.5.1	The Boyce-Codd Normal Form	208
6.5.2	The Third Normal Form	210
6.6	Properties of Decompositions	211
6.6.1	Lossless and Lossy Decompositions	212
6.6.2	Dependency-Preserving Decompositions	215
6.7	An Algorithm for BCNF Decomposition	219
6.8	Synthesis of 3NF Schemas	221
6.8.1	Minimal Cover	222
6.8.2	3NF Decomposition through Schema Synthesis	224
6.8.3	BCNF Decomposition through 3NF Synthesis	226
6.9	The Fourth Normal Form	228

- 6.10 Advanced 4NF Design* 233
 - 6.10.1 MVDs and Their Properties 234
 - 6.10.2 The Difficulty of Designing for 4NF 235
 - 6.10.3 A 4NF Decomposition How-To 238
- 6.11 Summary of Normal Form Decomposition 240
- Ⓢ 6.12 Case Study: Schema Refinement for the Student Registration System 241
- 6.13 Tuning Issues: To Decompose or Not to Decompose? 244
 - Bibliographic Notes 245
 - Exercises 246

7 Triggers and Active Databases 251

- 7.1 What Is a Trigger? 251
- 7.2 Semantic Issues in Trigger Handling 252
- 7.3 Triggers in SQL:1999 256
- 7.4 Avoiding a Chain Reaction 264
 - Bibliographic Notes 265
 - Exercises 265

8 Using SQL in an Application 267

- 8.1 What Are the Issues Involved? 267
- 8.2 Embedded SQL 268
 - 8.2.1 Status Processing 271
 - 8.2.2 Sessions, Connections, and Transactions 273
 - 8.2.3 Executing Transactions 274
 - 8.2.4 Cursors 276
 - 8.2.5 Stored Procedures on the Server 282
- 8.3 More on Integrity Constraints 285
- 8.4 Dynamic SQL 286
 - 8.4.1 Statement Preparation in Dynamic SQL 287
 - 8.4.2 Prepared Statements and the Descriptor Area* 290
 - 8.4.3 Cursors 293
 - 8.4.4 Stored Procedures on the Server 293
- 8.5 JDBC and SQLJ 294
 - 8.5.1 JDBC Basics 294
 - 8.5.2 Prepared Statements 297
 - 8.5.3 Result Sets and Cursors 297
 - 8.5.4 Obtaining Information about a Result Set 300
 - 8.5.5 Status Processing 300

8.5.6	Executing Transactions	301
8.5.7	Stored Procedures on the Server	302
8.5.8	An Example	303
8.5.9	SQLJ: Statement-Level Interface to Java	303
8.6	ODBC*	307
8.6.1	Prepared Statements	309
8.6.2	Cursors	309
8.6.3	Status Processing	312
8.6.4	Executing Transactions	312
8.6.5	Stored Procedures on the Server	313
8.6.6	An Example	313
8.7	Comparison	315
	Bibliographic Notes	316
	Exercises	316

PART THREE Optimizing DBMS Performance and Transaction Processing **319**

9 Physical Data Organization and Indexing **321**

9.1	Disk Organization	322
9.1.1	RAID Systems	326
9.2	Heap Files	329
9.3	Sorted Files	333
9.4	Indices	337
9.4.1	Clustered versus Unclustered Indices	340
9.4.2	Sparse versus Dense Indices	342
9.4.3	Search Keys Containing Multiple Attributes	344
9.5	Multilevel Indexing	347
9.5.1	Index-Sequential Access	350
9.5.2	B ⁺ Trees	353
9.6	Hash Indexing	360
9.6.1	Static Hashing	360
9.6.2	Dynamic Hashing Algorithms	363
9.7	Special-Purpose Indices	371
9.7.1	Bitmap Indices	371
9.7.2	Join Indices	372
9.8	Tuning Issues: Choosing Indices for an Application	373
	Bibliographic Notes	374
	Exercises	375

10	The Basics of Query Processing	379
10.1	Overview of Query Processing	379
10.2	External Sorting	380
10.3	Computing Projection, Union, and Set Difference	384
10.4	Computing Selection	386
10.4.1	Selections with Simple Conditions	387
10.4.2	Access Paths	389
10.4.3	Selections with Complex Conditions	391
10.5	Computing Joins	392
10.5.1	Computing Joins Using Simple Nested Loops	393
10.5.2	Sort-Merge Join	396
10.5.3	Hash Join	398
10.6	Multirelational Joins*	399
10.7	Computing Aggregate Functions	401
	Bibliographic Notes	401
	Exercises	401
11	An Overview of Query Optimization	405
11.1	Query Processing Architecture	405
11.2	Heuristic Optimization Based on Algebraic Equivalences	407
11.3	Estimating the Cost of a Query Execution Plan	410
11.4	Estimating the Size of the Output	418
11.5	Choosing a Plan	420
	Bibliographic Notes	425
	Exercises	425
12	Database Tuning	429
12.1	Disk Caches	430
12.1.1	Tuning the Cache	431
12.2	Tuning the Schema	433
12.2.1	Indices	433
12.2.2	Denormalization	440
12.2.3	Repeating Groups	441
12.2.4	Partitioning	442
12.3	Tuning the Data Manipulation Language	443
12.4	Tools	446
12.5	Managing Physical Resources	447
12.6	Influencing the Optimizer	448
	Bibliographic Notes	451

Exercises 451

13 An Overview of Transaction Processing 455

- 13.1 Isolation 455
 - 13.1.1 Serializability 456
 - 13.1.2 Two-Phase Locking 458
 - 13.1.3 Deadlock 462
 - 13.1.4 Locking in Relational Databases 463
 - 13.1.5 Isolation Levels 465
 - 13.1.6 Lock Granularity and Intention Locks 468
 - 13.1.7 Summary 471
- 13.2 Atomicity and Durability 472
 - 13.2.1 The Write-Ahead Log 472
 - 13.2.2 Recovery from Mass Storage Failure 476
- 13.3 Implementing Distributed Transactions 477
 - 13.3.1 Atomicity and Durability—The Two-Phase Commit Protocol 478
 - 13.3.2 Global Serializability and Deadlock 480
 - 13.3.3 Replication 482
 - 13.3.4 Summary 484
 - Bibliographic Notes 484
 - Exercises 485

PART FOUR Software Engineering Issues and Documentation 487

14 Requirements and Specifications 489

- 14.1 Software Engineering Methodology 489
 - 14.1.1 UML Use Cases 490
- ④ 14.2 The Requirements Document for the Student Registration System 493
- ④ 14.3 Requirements Analysis—New Issues 500
- ④ 14.4 Specifying the Student Registration System 502
 - 14.4.1 UML Sequence Diagrams 503
- ④ 14.5 The Specification Document for the Student Registration System: Section III 504
- 14.6 The Next Step in the Software Engineering Process 506
 - Bibliographic Notes 506
 - Exercises 507